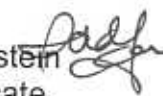




DEPARTMENT OF THE TREASURY  
INTERNAL REVENUE SERVICE  
WASHINGTON, D.C. 20224

DATE: November 14, 2003

MEMORANDUM FOR SUSAN L. SMOTER  
EXECUTIVE DIRECTOR,  
INTERNET DEVELOPMENT SERVICES

FROM: Maya A. Bernstein   
Privacy Advocate

SUBJECT: Privacy Impact Assessment for IRS.GOV  
Upgrade Utilities

The Office of the Privacy Advocate has reviewed the Privacy Impact Assessment for the IRS.GOV Upgrade Utilities. Based on the information you provided, we do not have any privacy concerns that would preclude this system from operating. However, a revised PIA is required when considering any future upgrades or major modifications to the system.

We will forward a copy of the PIA to the Director, Modernization and System Security to be included in the Security Accreditation Package for formal acceptance for operation. If you have any questions, please contact me at 202-927-5170 or Susan Dennis at 202-622-5438.

Attachment

cc: Director, Information Technology Services Security and Certification

Date November 10, 2003

MEMORANDUM FOR MAYA A. BERNSTEIN  
PRIVACY ADVOCATE

FROM: SUSAN L. SMOTER  
EXECUTIVE DIRECTOR  
INTERNET DEVELOPMENT SERVICES

SUBJECT: Request for Privacy Impact Assessment (PIA) –  
IRS.GOV Public Website

Purpose of the System:

The purpose of this submission is to assess the impact of new data being collected due to the upgrade of the WebTrends Reporting Center software, from WebTrends Version 4.0 E-Business Edition to WebTrends Version 6.0 Enterprise Edition. It also assesses the impact of new data being collect due to the upgrade of the Vignette Content Management Application (CMA), from Vignette CMA Version 6 to Vignette CMA Version 7.

The WebTrends Reporting Center software is used by designated IRS employees and contractors to measure and analyze the website performance of the IRS' public portal ([www.irs.gov](http://www.irs.gov), also known as IRS.GOV). The Vignette Content Management Application is used by designated IRS employees and contractors to create, update, and post the business areas' public web pages to IRS.Gov.

The IRS.gov website is intended to provide a wide variety of users with "self service" access to the information and resources they need from the IRS. To the extent that people can easily find answers to their questions on IRS.gov, they will not need to contact IRS, thereby reducing costs to the IRS. Besides providing many of the IRS's forms and publications as downloadable PDF (Portable Document Format) files, the site provides a great deal of information in the form of HTML (HyperText Mark-up Language) web pages that can be found by browsing or searching the site. The site also provides tools such as a withholding calculator, specially designed educational material such as "Tax Interactive," tax news, FAQs (Frequently Asked Questions with answers), and links to related information on non-IRS websites.

Neither tax return data nor Sensitive But Unclassified (SBU) data are accessible or otherwise used on the public portal. However, the IRS.Gov website provides links to online applications residing on the Registered User Portal (RUP). By utilizing proper identification, the RUP applications allow inquiries regarding claimed credits, refund or return status, or other information regarding taxpayer accounts. It should be noted that the RUP is not part of the IRS Public Portal,

and therefore is not covered by this PIA. A separate PIA for the RUP was approved in September 2003.

Name of Request Contact:

Name: Andrew J. LeBold  
Organization Name & Symbols: Internet Development Services  
OS:CIO:I:ET:ID:PP  
Mailing Address: NCFB, 5000 Ellin Road, Lanham MD 20706  
Phone Number (with area code): 202-283-0389

Name of Business System Owner:

Name: Susan L. Smoter  
Organization Name & Symbols: Internet Development Services  
OS:CIO:I:ET:ID  
Mailing Address: NCFB, 5000 Ellin Road, Lanham MD 20706  
Phone Number (with area code): 202-283-4881

Requested Operational Date: November 12, 2003

Category: *(Reason PIA is required--enter "y" or "n" and applicable dates)*

New System: n

Recertification? (if no change, enter date of last certification) n

Modification of existing system: y

Is this a National Standard Application (NSA): n

Is this a Modernization Project or System? n

If yes, the current milestone: Current Release

System of Records Number(s) (SORN):

Treasury/IRS Audit Trail and Security Records System 34.037

Attachment: PIA

## Data in the System

1. Describe the information (data elements and fields) available in the system in the following categories:

- A. Taxpayer
- B. Employee
- C. Audit Trail Information (including employee log-in info)
- D. Other (Describe)

A. All visitors to the website (i.e., Taxpayer, Employee, and others) will have their IP addresses logged and stored to analyze aggregate user behavior and to detect security violations. Additionally, standard access information will be logged and stored temporarily on Web, Application, Firewall, IDS (Intrusion Detection System), and database systems. The logged and stored information captured is crucial for security investigations but will not be used to uniquely identify users between/across sessions of web usages. The reporting software utilized will not collect any other taxpayer information.

B (1). The Vignette Content Management Application (CMA) of IRS.gov stores "Personnel Content Type" data and "User Profile" data on IRS employees who use the system. This system also distributes "CMA Administrative Reports" to designated IRS users. Outside the CMA system, password-related "Shared Secrets/Shared Secret Answers" are stored to maintain the business process for retrieving passwords. The details are provided below:

- CMA "Personnel Content Type" Data: IRS employees who either use the CMA, or provide content to an IRS employee who uses the CMA of IRS.gov, have their first name/first initial, middle name/middle initial, last name, suffix (e.g., Jr., Sr.), organization, phone number, and IRS email stored in the system as a "Personnel Content Type". As a CMA user creates a new content item for IRS.gov, he/she will identify a Content Maintainer (i.e., the person responsible for maintaining the content), and optionally identify a Content Provider. The Content Maintainer and Content Provider fields utilize the Personnel Content Type, which stores the IRS employee information described in the fields listed above.
- CMA "User Profile" Data: IRS employees who use the Vignette CMA of IRS.gov have all the information captured in the Personnel Content Type as part of their user profile. In addition, the system captures user name, password, roles (creator, reviewer, publisher), and project folder access (content areas) stored as part of their CMA "User Profile". This information is necessary to identify employees responsible for creating or modifying web pages to ensure that the information remains accurate and timely, to establish a formal span of control of published web material, and to maintain an audit trail.
- CMA "Administrative Reporting": The CMA provides reporting capabilities as a means to verify the currentness of content on the website as well as to monitor potential "broken links". The reporting also includes information on CMA system usage. There are three types of reports defined for CMA Release 2.0:
  - o Orphan Content Report – This report provides a list of web content that has no reference to it by any other web page or content item. This report is generated and distributed to the IRS Public Portal Branch (PPB) with information on who last modified the content. The orphan page report is then distributed to the CMA "Super Users" by the PPB.
  - o Unpublished and Expired Content Report – This report provides a list of unpublished content items, content items scheduled to expire within the next 60 days, and expired content items. This report is designed to capture Last Modified By (Full Name), Organization, and status/state information on content. This report is intended to be distributed to the PPB and Super Users.
  - o User Profile Report - This report provides high level CMA user activity information (e.g., Name, Organization, Login Activity). This report is designed to capture Full Name, Organization, Successful Logins, and Last Logon Date. The User Profile Report is only provided to the PPB.

- CMA "Shared Secrets/Shared Secret Answers": IRS employees who use the CMA of IRS.gov have a "shared secret question" and a "shared secret answer" captured and stored in a password protected file outside the CMA. This information will be used by the IRS.gov Support Line for user authentication in the event an IRS CMA user forgets his or her password.

B (2). IRS employees who use the WebTrends Reporting Center have their User ID and password stored as part of their user profile. This information is needed for logon purposes.

C. The CMA stores the User ID whenever a user creates or modifies any website content in the system.

D. On pages where website visitors voluntarily request information, publications, refund status, or other information, an appropriate application-specific privacy statement is posted. Each statement informs the visitor of the information being requested; why it is being requested; how it will be used and maintained; and, the impact if the information requested is not provided. Each page of IRS.gov provides a link to the IRS Web Privacy Policy as well as links to taxpayers' rights under the Privacy Act and other privacy protection statutes.

2. Describe/identify which data elements are obtained from files, databases, individuals, or any other sources.

- A. IRS
- B. Taxpayer
- C. Employee
- D. Other Federal Agencies (List agency)
- E. State and Local Agencies (List agency)
- F. Other third party sources (Describe)

The information provided is captured directly from the network traffic and passed to the IRS.gov infrastructure. The data, derived from visitor's activities on IRS.gov, will also be captured anonymously using JavaScript tagging on SmartSource Data Collectors (SDCs). The SDCs will be located at both Denver and Sterling sites. Previously, this visitor data was collected directly from the network traffic. All individual visitor information obtained from the Internet will be collected on the separate Registered User Portal, and will not be associated with this project. No additional information on individuals will be collected.

- a. What IRS files and databases are used?  
None
- b. What Federal Agencies are providing data for use in the system?  
None
- c. What State and Local Agencies are providing data for use in the system?  
None
- d. From what other third party sources will data be collected?  
None
- e. What information will be collected from the taxpayer/employee?

The following information will be collected from web pages accessed by visitors:

- IP Address of the visitor
- Date and Time (with Time Zone) of a page view
- Request details (web page)
- TCP/IP Packet details (in the case of IDS for tracking malicious packet attacks)
- Operating System of user
- Browser version
- Domain name
- Referring website
- Session cookie ID (used only for compiling subsequent page views into a unique, but anonymous, user session)
- Visitor Display Color Depth (e.g., 32 bits, 24 bits, etc.)



- Visitor Screen Resolution (e.g., 1024 x 768 pixels, 800 x 600 pixels, etc.)
- Number of Bytes received

Collection and analysis of this information in the aggregate will enable us to enhance site performance, make design improvements, improve informational materials available on our website, and improve customer service overall.

3. Is each data item required for the business purpose of the system? Yes. Explain.

- a. Will the system derive new data or create previously unavailable data about an individual through aggregation from the information collected?

Yes, but none of this information is personally identifiable data..

- b. Will the new data be placed in the individual's record (taxpayer or employee)? No.

- c. Can the system make determinations about taxpayers or employees that would not be possible without the new data?

Yes, the new system will provide the capacity for the IRS to determine the following aggregate user behavior (but not at an individual or group identifiable level).

- Determine the number of popular paths to and from IRS's website
- Enable the IRS to conduct scenario analysis from which the IRS can define scenarios and track conversion performance
- Enable the IRS to create drill down reports to obtain more detail information about the contents of a category
- Determine the geographic distribution of the visitors to the website

The system will also be able to determine whether and when an employee has created or modified website content in the CMA.

- d. How will the new data be verified for relevance and accuracy?

Accenture will only check the accuracy of data if the data is used in a forensic capacity.

4. How will each data item be verified for accuracy, timeliness, and completeness?

Accenture will not perform any proactive verification. In the event of a security breach, information will be verified through standard chain-of-evidence procedures and forensic methods by IRS or enforcement personnel.

5. Is there another source for the data? Explain how that source is or is not used.

No. Detailed web statistics and analyses are only provided by the WebTrends Reporting Center.

6. Generally, how will data be retrieved by the user?

Website visitors will not be able to retrieve data stored in web server, firewall, or network logs. In addition, website visitors will not be able to retrieve data stored by WebTrends or the WebTrends SmartSource Data Collection (SDC) servers. Designated IRS users will be able to retrieve web metric data via WebTrends software. All of this data collected is statistical in nature and retrievable at an aggregated-level by topic (e.g., hits, page views, length of visit, etc.) via the WebTrends Reporting Center.

IRS employees with access to the CMA may create or update only those web pages assigned to their business operating division.

Both WebTrends and CMA users have the ability to change their passwords.

7. Is the data retrievable by a personal identifier such as name, SSN, or other unique identifier? No.

### Access to the Data

8. Who will have access to the data in the system (Users, Managers, System Administrators, Developers, Others)?

Treasury cleared system administrators and the Senior Security Manager will have access to the raw data.

The WebTrends reports, generated from the data collected by the SDCs without any unique identifiers (e.g., IP address), will be provided to various IRS executives and employees through an Intranet website. In addition, approved IRS stakeholders will have access to WebTrends reports generated by data that is relevant to their assigned group. The exact list of approved personnel and specific access permissions is not yet defined and is likely to change as the needs of the business change. For more information about access controls, see response provided to question #9 below.

9. How is access to the data by a user determined and by whom?

Access to the raw data will be managed through the use of mandatory access controls on the systems and physical locks on the off-site backup storage. Log access will only be granted to the Systems Administrators and Senior Security Manager who will be assigned read-only access.

Access to aggregate data from WebTrends will be presented to IRS users with a valid username/password through a reporting web server. The username/password will be issued to approved IRS stakeholders and executives. Accenture will set profiles for different groups within the IRS and users will only be granted access to reports generated by data that is relevant to their assigned group.

Access to aggregate data from the CMA will be presented to IRS users with a valid username/password through a content management application. The username/password will be issued to approved IRS employees. Accenture will set profiles for different groups within the IRS, and users will only be granted access to content that is relevant to their assigned group.

10. Do other IRS systems provide, receive, or share data in the system? If YES, list the system(s) and describe which data is shared. If NO, continue to Question 12. No.

11. Have the IRS systems described in Item 10 received an approved Security Certification and Privacy Impact Assessment? N/A

12. Will other agencies provide, receive, or share data in any form with this system?

- a. Will other agencies share data or have access to data in this system (International, Federal, State, Local, Other)?

No other agencies will share or have access to the data maintained in the system except for required oversight or law enforcement purposes.

- b. How will the data be used by the agency?

The IRS will use the captured data to make enhancements to the website and to analyze the effectiveness of the website. Additionally, the captured information will be used by IRS executives and stakeholders for performance metrics and capacity planning.

In the event of a security issue, the captured information will be used as evidence by the appropriate law enforcement agencies.

- c. Who is responsible for assuring proper use of the data?

By contract, Qwest, Accenture, and the IRS have joint responsibility for assuring the proper use of the captured data.

- d. How will the system ensure that agencies only get the information they are entitled to under IRC 6103?

No. IRC§6103 data is available on the IRS. public website, nor is data shared with other agencies. Accenture will only provide reporting access via WebTrends to authorized IRS executives and stakeholders. Accenture will purge old IRS user accounts from the system before the installation of the new software.

On an annual basis, Accenture will purge unused IRS user accounts from the system, and a new Form 5081 will be required for renewal of user accounts.

## Administrative Controls of Data

13. What are the procedures for eliminating the data at the end of the retention period?

Raw data (web access logs) and system access logs are retained for one year. To enable the IRS to compare the statistical data from year-to-year, the WebTrends database will be retained up to four years.

The raw data will be stored on tapes at the end of four years and removed from the hard drives.

Any data that is retained for any significant duration is collected for web analysis and is not used to identify any individuals.

14. Will this system use technology in a new way? If "YES" describe. If "NO" go to Question 15.

Yes. The combination of client-side JavaScript and session cookies will be used on all IRS.gov web pages, as well as by some of the applications deployed on the website, to accurately analyze how visitors navigate through the IRS website at an aggregated level.

15. Will this system be used to identify or locate individuals or groups? If so, describe the business purpose for this capability.

Yes, but only for special circumstances. In the event of an attack, this system could provide data, which when combined with other forensic information, may contribute to the capability to identify or locate an individual. It should be noted that the data in the system alone is not sufficient to locate an individual or group. Identification of an individual or group will only be possible when the data in the system is combined with other data, such as access log information provided by an Internet Service Provider (ISP). This would be a law enforcement action accomplished via a system external to the IRS.gov website, and therefore outside the purview of this PIA.

The capability to identify, locate, and monitor groups of people will be provided for the purpose of gathering aggregate information for web statistical reports. No individual demographics or personally identifiable information is collected or monitored.

16. Will this system provide the capability to monitor individuals or groups? If yes, describe the business purpose for this capability and the controls established to prevent unauthorized monitoring.

Yes, but only for the purposes of aggregate information gathering for web statistics reports. The capability to identify, locate, and monitor groups of people will be provided for the purpose of gathering aggregate information for web statistical reports. Individual demographics (e.g., age, sex, religion, etc.) will not be collected or monitored.

17. Can use of the system allow IRS to treat taxpayers, employees, or others, differently? Explain.



The IRS.gov public website, WebTrends, and the CMA tool are not capable of influencing the treatment of groups or individuals, or making determinations regarding any actions by, or against groups or individuals.

Access to the CMA and WebTrends systems is restricted to IRS employees and IRS contractors based on their current job requirements. Additionally, in the event of an attack on the [www.irs.gov](http://www.irs.gov) infrastructure, the responsible IRS contractor may disable individual IP addresses from accessing the system in an effort to protect the website and to limit service degradation to other users.

18. Does the system ensure "due process" by allowing affected parties to respond to any negative determination, prior to final action?

The IRS.gov website only provides information to the public. The system is incapable of making determinations of any kind or of influencing legal actions that may be taken against any parties.

19. If the system is web-based, does it use persistent cookies or other tracking devices to identify web visitors?

No. The WebTrends system uses transient session cookies on all web pages to anonymously track users. These session cookies are non-persistent, and they stored in temporary memory. They are erased when the visitor's browser is closed. Session cookies allow the IRS to accurately analyze how visitors navigate through the IRS website at an aggregated level. Session cookies will also allow the IRS to perform site improvements based on the way that the website is actually used by visitors.

In the CMA, the internal IRS user has the option of using a persistent cookie to store their CMA User ID. This allows their User ID to be "auto-filled" in their browser, so that the IRS CMA user only types their password to login. This enables customization to accommodate personal preferences and is voluntarily enabled by the user.